

Split optimal policy iteration for LQR problems

Péter Koltai

April 22, 2014

Abstract

This technical report is concerned with the convergence properties of what we call the *split optimal policy iteration* for coupled LQR problems; see section 3.1 below. Interestingly, the iteration shows different convergence behavior for continuous and discrete time systems: while global convergence holds for both cases, we have local quadratic convergence for the continuous time case, but only linear convergence for the discrete time case—even though quadratic convergence is retained in the limit as the coupling between the subsystems vanishes.

Contents

1	Introduction	1
2	LQR problems: notation	3
2.1	Continuous time	3
2.2	Discrete time	3
3	Split optimal policy iteration for LQR problems	4
3.1	Two subsystems	4
3.2	Arbitrary number of subsystems	4
4	Convergence: continuous time systems	5
5	Convergence: discrete time systems	9

1 Introduction

We consider an iterative solution method for optimal control problems for time-invariant coupled systems. For time-invariant (autonomous) systems the optimal controller, if it exists, is given by a *feedback* (also called *policy*), i.e. a mapping from the state into the admissible control space. We consider the solution to the optimal control problem being equal to finding this policy. The system structure considered here is such that each subsystem is characterized by its own state and control variables, and the evolution of the state variables is coupled through the dynamics. For such systems one could attempt to solve the optimal control problem by optimizing the policy of just one subsystem while keeping the policies of the other subsystems fixed, then doing the same for the next subsystem, and so on. Since in each step the optimal policy with respect to the current policies of the other subsystems is chosen by the

current subsystem, we call this split optimal policy iteration. The process is exemplified for LQR problems with two subsystems in section 3.1.

The purpose of this technical report is the *qualitative and quantitative* analysis of the split optimal policy iteration applied to continuous and discrete time LQR problems. The motivation is rather phenomenological, since it is not immediate to the author whether this procedure has any numerical advantage over solving the algebraic Riccati equation of the full problem directly. However, if the here observed *fast* convergence for weakly coupled systems carries over to nonlinear systems as well, then this could open pathways for the design of centralized (sub-)optimal controllers for general nonlinear systems, as discussed in [KJ14b].

The idea of split iterative optimization is by no means new. For instance, iterative optimization over reduced variables can be considered for minimizing scalar functions $f : \mathbb{R}^n \rightarrow \mathbb{R}$ whenever the structure of f is such that it is easier to solve $\min_{x_i} f(x_1, \dots, x_n)$ for any $i = 1, \dots, n$ than the full problem, and the cyclical repetition of this procedure is expected to converge to a reasonable solution [BH02]. Note the difference to the setting considered here: in optimal control one optimizes the optimal value function in every state of the state space, hence we are dealing with multiple objective functions. However, Bellman’s optimality principle [Bel57] states that these objectives are not concurrent; in order to be optimal, every end-part of a path has to be optimal itself. This would suggest that instruments from this field are likely to be able to be carried over to our setting. Indeed, the monotonicity in Lemma 2 can be seen as an example for this.

Another property of the iterative procedure described here is that for general nonlinear systems it is not guaranteed to converge to the optimal solution—just as the iteration applied to minimize $f : \mathbb{R}^n \rightarrow \mathbb{R}$ can converge to a local but not global minimum. However, a fixed point of the iteration has a game-theoretical interpretation: it is a *Nash equilibrium* [OR94]. In a Nash equilibrium the policy of each player (here: the subsystem controller) is optimal with respect to the policies of the other players.¹ In game-theoretical terms our procedure is a *best-response strategy* iteration, or a kind of *fictitious play*, which has been proposed to compute a Nash equilibrium of a game [Bro51, Rob51, MS96b, MS96a]. It has to be noted, that games often only possess *mixed* Nash equilibria, which are randomized strategies. Optimal control problems (even with stochastic components, as *Markov decision processes*), however, have deterministic optimal policies. But coupled optimal control problems can be rewritten as static (collaborative) games, see e.g. the lecture notes [Joh]. It is not clear to the author whether, and to which extent, does this connection allow to carry over the convergence results from game theory to the present setting.

As the aim of the subsystems is to optimally accomplish a joint goal, it is not surprising that very similar ideas occur in *multiagent systems* and *reinforcement learning* too [Lit01, MHeK⁺98, LR00, GKP01]. Certainly, the idea of separability can be found in the *dynamic programming* literature too [Ber07].

To facilitate the navigation in this report, the main theorems are framed and all proofs have a gray background (and can be skipped if one is only interested in the main results).

¹The reader might be irritated by the fact that in our coupled optimal control setting the subsystems follow a joint *cooperative* goal, while game theory usually considers *competitive* situations. However, the “definition” of a Nash equilibrium given here does not include the players’ intentions, it only assumes they are optimal with respect to each other. This covers both situations.

2 LQR problems: notation

2.1 Continuous time

Linear-quadratic regulator (LQR) problems arise in the context when the linear system

$$\dot{x}(t) = \mathbf{A}x(t) + \mathbf{B}u(t) \quad (1)$$

shall be controlled to the origin in an optimal way. Here, $\mathbf{A} \in \mathbb{R}^{m \times m}$ denotes the system matrix and $\mathbf{B} \in \mathbb{R}^{m \times r}$ denotes the input matrix. Further, let $\mathbf{Q} \in \mathbb{R}^{m \times m}$ and $\mathbf{R} \in \mathbb{R}^{r \times r}$ be symmetric positive definite matrices. The control task is to find a function $u : [0, \infty) \rightarrow \mathcal{U}$, generating a trajectory $\{x(t)\}_{t \geq 0}$ through (1), such that the accumulated costs

$$\int_0^\infty x(t)^T \mathbf{Q} x(t) + u(t)^T \mathbf{R} u(t) dt$$

are minimal. By construction, this already implies that $x(t) \rightarrow 0$ as $t \rightarrow \infty$. It turns out that the optimality principle is equivalent to the *continuous algebraic Riccati equation*

$$\mathbf{A}^T \mathbf{P} + \mathbf{P} \mathbf{A} - \mathbf{P} \mathbf{B} \mathbf{R}^{-1} \mathbf{B}^T \mathbf{P} + \mathbf{Q} = \mathbf{0} \quad (2)$$

where the unique symmetric positive definite solution \mathbf{P}^{opt} of this equation yields the optimal value function $V(x) = x^T \mathbf{P}^{\text{opt}} x$. Moreover, the optimal feedback is given by

$$\mu(x) = \mathbf{F}^{\text{opt}} x = -\mathbf{R}^{-1} \mathbf{B}^T \mathbf{P}^{\text{opt}} x. \quad (3)$$

2.2 Discrete time

Linear-quadratic regulator problems arise in the context when the linear system

$$x(k+1) = \mathbf{A}x(k) + \mathbf{B}u(k) \quad (4)$$

shall be controlled to the origin in an optimal way. Here, $\mathbf{A} \in \mathbb{R}^{m \times m}$ denotes the system matrix and $\mathbf{B} \in \mathbb{R}^{m \times r}$ denotes the input matrix. Further, let $\mathbf{Q} \in \mathbb{R}^{m \times m}$ and $\mathbf{R} \in \mathbb{R}^{r \times r}$ be symmetric positive definite matrices. The control task is to find a sequence $\{u(k)\}_{k \geq 0}$, generating a sequence $\{x(k)\}_{k \geq 0}$ through (4), such that the accumulated costs

$$\sum_{k \geq 0} x(k)^T \mathbf{Q} x(k) + u(k)^T \mathbf{R} u(k)$$

are minimal. By construction, this already implies that $x(k) \rightarrow 0$ as $k \rightarrow \infty$. It turns out that the optimality principle is equivalent to the *discrete algebraic Riccati equation*

$$\mathbf{P} = \mathbf{A}^T \mathbf{P} \mathbf{A} - \mathbf{A}^T \mathbf{P} \mathbf{B} \left(\mathbf{R} + \mathbf{B}^T \mathbf{P} \mathbf{B} \right)^{-1} \mathbf{B}^T \mathbf{P} \mathbf{A} + \mathbf{Q}, \quad (5)$$

where the unique symmetric positive definite solution \mathbf{P}^{opt} of this equation yields the optimal value function $V(x) = x^T \mathbf{P}^{\text{opt}} x$. Moreover, the optimal feedback is given by

$$\mu(x) = \mathbf{F}^{\text{opt}} x = - \left(\mathbf{R} + \mathbf{B}^T \mathbf{P} \mathbf{B} \right)^{-1} \mathbf{B}^T \mathbf{P}^{\text{opt}} \mathbf{A} x. \quad (6)$$

3 Split optimal policy iteration for LQR problems

3.1 Two subsystems

In order to get a better intuition about how the split optimal policy iteration works, we show an update step for the case of two subsystems. For this, let us partition the involved matrices \mathbf{A} , \mathbf{B} , \mathbf{Q} , and \mathbf{R} into blocks according to the subsystem decomposition, i.e.

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix},$$

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_2 \end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} \mathbf{R}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_2 \end{pmatrix}.$$

Assuming that some feedback matrix

$$\mathbf{F} = \begin{pmatrix} \mathbf{F}_{11} & \mathbf{F}_{12} \\ \mathbf{F}_{21} & \mathbf{F}_{22} \end{pmatrix}$$

is given (partitioned the same way), we now show how the feedback ν_1 of the first subsystem is updated by the split optimal policy iteration algorithm. The update works analogously for the second subsystem as well. First, note that $\nu_1(x) = (\mathbf{F}_{11} \ \mathbf{F}_{12})x$ and $\nu_2(x) = (\mathbf{F}_{21} \ \mathbf{F}_{22})x$. Since ν_2 is fixed during this update step, we have to merge it into the matrices \mathbf{A} and \mathbf{Q} . We obtain a new system matrix

$$\mathbf{A}^{(1)} = \mathbf{A} + \begin{pmatrix} \mathbf{B}_{12} \\ \mathbf{B}_{22} \end{pmatrix} (\mathbf{F}_{21} \ \mathbf{F}_{22}),$$

and a new state cost matrix

$$\mathbf{Q}^{(1)} = \mathbf{Q} + (\mathbf{F}_{21} \ \mathbf{F}_{22})^T \mathbf{R}_2 (\mathbf{F}_{21} \ \mathbf{F}_{22}).$$

The control- and control cost matrices reduce to

$$\mathbf{B}^{(1)} = \begin{pmatrix} \mathbf{B}_{11} \\ \mathbf{B}_{21} \end{pmatrix}, \text{ and } \mathbf{R}^{(1)} = \mathbf{R}_1.$$

Moreover, we obtain an LQR problem with matrix quadruple $\mathbf{A}^{(1)}$, $\mathbf{B}^{(1)}$, $\mathbf{Q}^{(1)}$, and $\mathbf{R}^{(1)}$. Solving this LQR problem via (2) and (3) or (5) and (6) yields $\nu_1^{\text{new}}(x) = (\mathbf{F}_{11}^{\text{new}} \ \mathbf{F}_{12}^{\text{new}})x$.

Now, that the new feedback matrix

$$\mathbf{F}^{\text{new}} = \begin{pmatrix} \mathbf{F}_{11}^{\text{new}} & \mathbf{F}_{12}^{\text{new}} \\ \mathbf{F}_{21} & \mathbf{F}_{22} \end{pmatrix}$$

is obtained, the iteration continues with the update of the 2nd subsystem with \mathbf{F}^{new} replacing \mathbf{F} .

3.2 Arbitrary number of subsystems

To generalize the approach described in the previous section, consider an LQR problem given by the matrices \mathbf{A} , \mathbf{B} , \mathbf{Q} , and \mathbf{R} , where \mathbf{Q} and \mathbf{R} are block diagonal symmetric positive definite.

Let Π_i , $i = 1, \dots, n$ with n being the number of subsystems, be a block column matrix consistent with the decomposition of \mathbf{R} , such that the i^{th} block entry is the identity matrix \mathbf{I} and all others are $\mathbf{0}$, i.e. Π_i is of the form $(\mathbf{0}, \dots, \mathbf{0}, \mathbf{I}, \mathbf{0}, \dots, \mathbf{0})^T$. Analogously as before, this yields

$$\begin{aligned} \mathbf{A}^{(i)} &= \mathbf{A} + \mathbf{B}(\mathbf{I} - \Pi_i \Pi_i^T) \mathbf{F}, \\ \mathbf{B}^{(i)} &= \mathbf{B} \Pi_i, \\ \mathbf{Q}^{(i)} &= \mathbf{Q} + \mathbf{F}^T (\mathbf{I} - \Pi_i \Pi_i^T) \mathbf{R} (\mathbf{I} - \Pi_i \Pi_i^T) \mathbf{F}, \\ \mathbf{R}^{(i)} &= \Pi_i^T \mathbf{R} \Pi_i. \end{aligned} \tag{7}$$

Solving the algebraic Riccati equations with these matrices gives \mathbf{P}^{new} , and the feedback matrix is updated by

$$\begin{aligned} \Pi_i^T \mathbf{F}^{\text{new}} &= \begin{cases} -\mathbf{R}^{(i)-1} \mathbf{B}^{(i)T} \mathbf{P}^{\text{new}}, & \text{continuous time} \\ -(\mathbf{R}^{(i)} + \mathbf{B}^{(i)T} \mathbf{P}^{\text{new}} \mathbf{B}^{(i)})^{-1} \mathbf{B}^{(i)T} \mathbf{P}^{\text{new}} \mathbf{A}^{(i)}, & \text{discrete time} \end{cases}, \\ (\mathbf{I} - \Pi_i^T) \mathbf{F}^{\text{new}} &= \mathbf{F}. \end{aligned}$$

4 Convergence: continuous time systems

We can split our considerations here to those concerning global, and those concerning local convergence properties. We start with global convergence.

Theorem 1 (Global convergence). *Let us start the split optimal policy iteration with the initial guess \mathbf{F}^0 and with updating the i^{th} subsystem first, where i is arbitrary. Then, if the pair $(\mathbf{A}^{(i)}, \mathbf{B}^{(i)})$ is controllable, then*

- (i) *all the iterates \mathbf{F}^k , $k \geq 1$, stabilize the global system;*
- (ii) *the value functions V^k corresponding to the \mathbf{F}^k satisfy $V^{k+1} \leq V^k$ pointwise; and*
- (iii) *$\mathbf{F}^k \rightarrow \mathbf{F}^{\text{opt}}$ as $k \rightarrow \infty$, where \mathbf{F}^{opt} is the optimal feedback matrix of the original problem.*

In order to prove this theorem, we break it down into several lemmas. Without mentioning it again, the controllability condition in the theorem is assumed to be valid throughout this section. First, we show the monotony of the value functions. Note that, since all the subproblems solved during the iteration are LQR problems themselves, $V^k(x) = x^T \mathbf{P}^k x$ for some symmetric positive definite \mathbf{P}^k .

Lemma 2. *We have $\mathbf{P}^{k+1} \preceq \mathbf{P}^k$ for $k \geq 0$, and \mathbf{F}^k is stabilizing for every $k \geq 1$. Here, $\mathbf{A} \preceq \mathbf{B}$ for two symmetric matrices if and only if $\mathbf{B} - \mathbf{A}$ is symmetric positive semidefinite.*

Proof. Even though the iteration is based on the successive solution of subproblems, they are constructed such that the value function they optimize is the same object in each step: it is $\int_0^\infty x(t)^T \mathbf{Q} x(t) + u(t)^T \mathbf{R} u(t) dt$ with $u(t) = \mathbf{F}^k x(t)$ for the k^{th} iterate. By not changing the \mathbf{F}^k in a step we would have $\mathbf{P}^{k+1} = \mathbf{P}^k$. But since in each step the optimal control (for the corresponding subproblem) is taken, we obtain $\mathbf{P}^{k+1} \preceq \mathbf{P}^k$ by optimality, and all corresponding feedbacks are stabilizing. The feedback matrix \mathbf{F}^1 is stabilizing by assumption. \square

Now, that we have a monotonically decreasing sequence of value functions, their convergence has to be established.

Lemma 3. *Let $\{\mathbf{P}^k\}_{k \geq 1}$ a sequence of symmetric positive definite matrices, such that $\mathbf{P}^{k+1} \preceq \mathbf{P}^k$ for every $k \geq 1$. Then $\mathbf{P}^k \rightarrow \bar{\mathbf{P}}$ as $k \rightarrow \infty$ for some symmetric positive semidefinite $\bar{\mathbf{P}}$.*

Proof. (This elegant proof is due to Benedict Dingfelder.) By assumption, $\{v^T \mathbf{P}^k v\}_{k \geq 1}$ is a monotonically decreasing sequence of positive real numbers for any $v \in \mathbb{R}^m$, hence it converges, say, to $p_v \geq 0$. Setting $v = e_i$, the i^{th} canonical vector, we obtain $\mathbf{P}_{ii}^k \rightarrow p_{e_i}$ as $k \rightarrow \infty$. With $v_{ij} := e_i + e_j$ we have $v_{ij}^T \mathbf{P}^k v_{ij} = \mathbf{P}_{ii}^k + \mathbf{P}_{jj}^k + 2\mathbf{P}_{ij}^k$, since all \mathbf{P}^k are symmetric. The left hand side of this equation converges, just as the diagonal entries of the matrix, shown above. It follows that \mathbf{P}_{ij} converges for every $i, j \in \{1, \dots, m\}$, hence \mathbf{P}^k converges element wise (and thus in any norm) to a matrix $\bar{\mathbf{P}}$. Since the set of positive semidefinite matrices is a closed subspace of $\mathbb{R}^{m \times m}$, the matrix $\bar{\mathbf{P}}$ is contained in this subspace, as the limit of symmetric positive definite matrices. \square

Next is to show that the feedback matrices converge to a fixed point of the split optimal policy iteration.

Lemma 4. *It holds $\mathbf{F}^k \rightarrow \bar{\mathbf{F}}$ as $k \rightarrow \infty$, where $\bar{\mathbf{F}} = -\mathbf{R}^{-1} \mathbf{B}^T \bar{\mathbf{P}}$ is a fixed point of the split optimal policy iteration.*

Proof. From the update assignment for the feedback matrices we have $\Pi_i^T \mathbf{F}^{k_i} = -\mathbf{R}^{(i)-1} \mathbf{B}^{(i)T} \mathbf{P}^{k_i}$, where the k_i are indices of the iteration steps where the i^{th} subsystem is updated. Thus, $\Pi_i^T \mathbf{F}^k \rightarrow -\mathbf{R}^{(i)-1} \mathbf{B}^{(i)T} \bar{\mathbf{P}}$ as $k \rightarrow \infty$, because $\mathbf{P}^{k_i} \rightarrow \bar{\mathbf{P}}$ as $k_i \rightarrow \infty$, independently from i . Since $\sum_i \Pi_i \Pi_i^T = \mathbf{I} \in \mathbb{R}^{m \times m}$, we have

$$\begin{aligned} \bar{\mathbf{F}} &= \sum_i \Pi_i \Pi_i^T \bar{\mathbf{F}} \\ &= -\sum_i \Pi_i (\Pi_i^T \mathbf{R} \Pi_i)^{-1} \Pi_i^T \mathbf{B}^T \bar{\mathbf{P}} \\ &= -\mathbf{R}^{-1} \mathbf{B}^T \bar{\mathbf{P}} \end{aligned}$$

where the last equality follows from \mathbf{R} being block diagonal.

In order to see that $\bar{\mathbf{F}}$ is a fixed point of the split optimal policy iteration, we investigate how \mathbf{F}^{k+1} is obtained from \mathbf{F}^k . The mapping $g^i : \mathbf{F}^k \mapsto \mathbf{F}^{k+1}$ can be decomposed into the following steps:

$$g^i : \mathbf{F}^k \mapsto (\mathbf{A}^{(i)}, \mathbf{B}^{(i)}, \mathbf{Q}^{(i)}, \mathbf{R}^{(i)}) \mapsto \mathbf{P}^{k+1} \mapsto \mathbf{F}^{k+1}. \quad (8)$$

The first and last mapping are obviously continuous (and arbitrarily often differentiable) by (7) and (3); the second one is even analytic [Del84]. Thus, g^i is continuous for every i , and the equation $\mathbf{F}^{k+1} = g^i(\mathbf{F}^k)$ yields $\bar{\mathbf{F}} = g^i(\bar{\mathbf{F}})$ as $k \rightarrow \infty$. \square

We are now ready to show the theorem. It remains to be shown that the fixed point $\bar{\mathbf{F}}$ of the split optimal policy iteration coincides with the optimal solution of the initial problem, i.e. with \mathbf{F}^{opt} .

Proof of Theorem 1. Statements (i) and (ii) are shown in Lemma 2. Statement (iii) follows from Lemma 4, if we show that any fixed point of the split optimal policy iteration (more precisely the associated matrix $\bar{\mathbf{P}}$) solves the continuous algebraic Riccati equation (2), and by uniqueness of its solutions [Son98] coincides with the optimal solution.

Since $\bar{\mathbf{P}}$ and $\bar{\mathbf{F}}$ are fixed points of the iteration, they solve (2) with the modified matrices (7) for every i . This reads as

$$\begin{aligned} \mathbf{0} = & \bar{\mathbf{P}}\mathbf{B}\Pi_i \left(\Pi_i^T \mathbf{R} \Pi_i \right)^{-1} \Pi_i^T \mathbf{B}^T \bar{\mathbf{P}} - \bar{\mathbf{P}} \left(\mathbf{A} + \mathbf{B} \left(\mathbf{I} - \Pi_i \Pi_i^T \right) \bar{\mathbf{F}} \right) \\ & - \left(\mathbf{A} + \mathbf{B} \left(\mathbf{I} - \Pi_i \Pi_i^T \right) \bar{\mathbf{F}} \right)^T \bar{\mathbf{P}} - \mathbf{Q} - \bar{\mathbf{F}}^T \left(\mathbf{I} - \Pi_i \Pi_i^T \right) \mathbf{R} \left(\mathbf{I} - \Pi_i \Pi_i^T \right) \bar{\mathbf{F}}. \end{aligned}$$

It is easy to see by the block diagonal structure of \mathbf{R} that

$$\Pi_i \left(\Pi_i^T \mathbf{R} \Pi_i \right)^{-1} \Pi_i^T = \Pi_i \Pi_i^T \mathbf{R}^{-1},$$

and

$$\left(\mathbf{I} - \Pi_i \Pi_i^T \right) \mathbf{R} \left(\mathbf{I} - \Pi_i \Pi_i^T \right) = \mathbf{R} \left(\mathbf{I} - \Pi_i \Pi_i^T \right) = \left(\mathbf{I} - \Pi_i \Pi_i^T \right) \mathbf{R}.$$

Using these identities and $\bar{\mathbf{F}} = -\mathbf{R}^{-1} \mathbf{B}^T \bar{\mathbf{P}}$ from Lemma 4, the above equation becomes

$$\begin{aligned} \mathbf{0} = & \bar{\mathbf{P}}\mathbf{B}\Pi_i \Pi_i^T \mathbf{R}^{-1} \mathbf{B}^T \bar{\mathbf{P}} - \bar{\mathbf{P}}\mathbf{A} - \mathbf{A}^T \bar{\mathbf{P}} \\ & + \bar{\mathbf{P}}\mathbf{B} \left(\mathbf{I} - \Pi_i \Pi_i^T \right) \mathbf{R}^{-1} \mathbf{B}^T \bar{\mathbf{P}} + \bar{\mathbf{P}}\mathbf{B}\mathbf{R}^{-1} \left(\mathbf{I} - \Pi_i \Pi_i^T \right) \mathbf{B}^T \bar{\mathbf{P}} \\ & - \mathbf{Q} - \bar{\mathbf{P}}\mathbf{B} \left(\mathbf{I} - \Pi_i \Pi_i^T \right) \mathbf{R}^{-1} \mathbf{B}^T \bar{\mathbf{P}} \end{aligned}$$

By collecting all terms starting with $\bar{\mathbf{P}}\mathbf{B}$ we obtain

$$\begin{aligned} \mathbf{0} = & \bar{\mathbf{P}}\mathbf{B} \left(\Pi_i \Pi_i^T \mathbf{R}^{-1} + \left(\mathbf{I} - \Pi_i \Pi_i^T \right) \mathbf{R}^{-1} + \mathbf{R}^{-1} \left(\mathbf{I} - \Pi_i \Pi_i^T \right) - \left(\mathbf{I} - \Pi_i \Pi_i^T \right) \mathbf{R}^{-1} \right) \mathbf{B}^T \bar{\mathbf{P}} \\ & - \bar{\mathbf{P}}\mathbf{A} - \mathbf{A}^T \bar{\mathbf{P}} - \mathbf{Q} \\ = & \bar{\mathbf{P}}\mathbf{B}\mathbf{R}^{-1} \mathbf{B}^T \bar{\mathbf{P}} - \bar{\mathbf{P}}\mathbf{A} - \mathbf{A}^T \bar{\mathbf{P}} - \mathbf{Q}, \end{aligned}$$

hence $\bar{\mathbf{P}}$ solves (2), and this proves the claim. \square

Remark 5. *It is enough to have controllability with respect to one subsystem: if there is a non-controllable subsystem, we just jump to the next without changing the feedback matrix as long as we find a subsystem with respect to which one has controllability.*

Locally, around the optimal solution, the iteration can be shown to converge quickly:

Theorem 6 (Local speed of convergence). *The split optimal policy iteration for continuous time LQR problem converges locally quadratically; i.e. for $\|\mathbf{F}^k - \mathbf{F}^{\text{opt}}\|$ sufficiently small we have*

$$\|\mathbf{F}^{k+n} - \mathbf{F}^{\text{opt}}\| \leq C \|\mathbf{F}^k - \mathbf{F}^{\text{opt}}\|^2$$

for some $C > 0$ which is independent of the \mathbf{F}^k .

Proof. The iteration is a consecutive application of the mappings g^i from (8) for i cyclically sweeping over the subsystem indices $\{1, \dots, n\}$. One cycle can be written as $g := g^{j-1} \circ g^{j-2} \circ \dots \circ g^{j+1} \circ g^j$ for some starting subsystem index j . The global convergence of the successive iterates of g to \mathbf{F}^{opt} from (3) was established above. Since $g(\mathbf{F}^{\text{opt}}) = \mathbf{F}^{\text{opt}}$, and g is arbitrary often differentiable, the local convergence depends on the spectral radius of the derivative of g at \mathbf{F}^{opt} .

We will show local quadratic convergence in the sense that

$$\|g(\mathbf{F}) - \mathbf{F}^{\text{opt}}\| \leq C \|\mathbf{F} - \mathbf{F}^{\text{opt}}\|^2$$

for any \mathbf{F} from a given neighborhood of the optimal solution, and some $C > 0$ independent of \mathbf{F} . Since g is a composition of the g^i , which all update different rows of \mathbf{F} during the iteration, it suffices to show quadratic convergence for the rows updated by a specific g^i for an arbitrary i , and the claim follows. Hence, fix $i \in \{1, \dots, n\}$ and consider the map g^i from (8). Quadratic convergence follows if $Dg^i(\mathbf{F}^{\text{opt}}) = 0$. Since the map $\mathbf{P}^{k+1} \mapsto \mathbf{F}^{k+1}$ is independent of \mathbf{F}^k , we have $Dg^i(\mathbf{F}^{\text{opt}}) = 0$ if $\partial \mathbf{P}^{k+1} / \partial \mathbf{F}^k(\mathbf{F}^{\text{opt}}) = 0$.

To see this, note that \mathbf{P}^{k+1} is the unique symmetric positive definite solution of $r^i(\mathbf{P}^{k+1}, \mathbf{F}^k) = 0$, where

$$r^i(\mathbf{P}, \mathbf{F}) := \mathbf{P}\mathbf{B}^{(i)}\mathbf{R}^{(i)-1}\mathbf{B}^{(i)\top}\mathbf{P} - \mathbf{P}\mathbf{A}^{(i)} - \mathbf{A}^{(i)\top}\mathbf{P} - \mathbf{Q}^{(i)},$$

with the definitions from (7), and the matrices $\mathbf{A}^{(i)}$ and $\mathbf{Q}^{(i)}$ depending on \mathbf{F} . By the implicit function theorem one has for $\mathbf{P}(\mathbf{F})$ satisfying $r^i(\mathbf{P}(\mathbf{F}), \mathbf{F}) = 0$ that

$$\frac{\partial \mathbf{P}}{\partial \mathbf{F}}(\mathbf{F}) = - \left(\frac{\partial r^i}{\partial \mathbf{P}}(\mathbf{P}, \mathbf{F}) \right)^{-1} \frac{\partial r^i}{\partial \mathbf{F}}(\mathbf{P}, \mathbf{F}),$$

in some neighborhood of any point such that $\partial r^i / \partial \mathbf{P}$ is invertible. We will show $\frac{\partial r^i}{\partial \mathbf{F}}(\mathbf{P}^{\text{opt}}, \mathbf{F}^{\text{opt}}) = 0$, implying the claim.

Differentiating r^i yields

$$\begin{aligned} \frac{\partial r^i}{\partial \mathbf{F}}(\mathbf{P}, \mathbf{F}) \cdot \Delta = & -\mathbf{P}\mathbf{B} \left(\mathbf{I} - \Pi_i \Pi_i^\top \right) \Delta - \mathbf{F}^\top \left(\mathbf{I} - \Pi_i \Pi_i^\top \right) \mathbf{R} \left(\mathbf{I} - \Pi_i \Pi_i^\top \right) \Delta \\ & - \Delta^\top \left(\mathbf{I} - \Pi_i \Pi_i^\top \right) \mathbf{B}^\top \mathbf{P} - \Delta^\top \left(\mathbf{I} - \Pi_i \Pi_i^\top \right) \mathbf{R} \left(\mathbf{I} - \Pi_i \Pi_i^\top \right) \mathbf{F}, \end{aligned} \quad (9)$$

for every $\Delta \in \mathbb{R}^{r \times m}$. Substituting $\mathbf{F}^{\text{opt}} = -\mathbf{R}^{-1}\mathbf{B}^T\mathbf{P}^{\text{opt}}$, we get

$$\begin{aligned}\mathbf{F}^{\text{opt}T} \left(\mathbf{I} - \Pi_i \Pi_i^T \right) \mathbf{R} \left(\mathbf{I} - \Pi_i \Pi_i^T \right) &= -\mathbf{P}^{\text{opt}} \mathbf{B} \mathbf{R}^{-1} \left(\mathbf{R} - \Pi_i \Pi_i^T \mathbf{R} \right) \left(\mathbf{I} - \Pi_i \Pi_i^T \right) \\ &= -\mathbf{P}^{\text{opt}} \mathbf{B} \left(\mathbf{I} - \Pi_i \Pi_i^T \right)^2 \\ &= -\mathbf{P}^{\text{opt}} \mathbf{B} \left(\mathbf{I} - \Pi_i \Pi_i^T \right),\end{aligned}$$

since $\mathbf{R}^{-1}\Pi_i \Pi_i^T \mathbf{R} = \Pi_i \Pi_i^T$ by the block diagonal structure of \mathbf{R} , and $(\Pi_i \Pi_i^T)^2 = \Pi_i \Pi_i^T$. Using this in (9) shows $\frac{\partial^j}{\partial \mathbf{F}}(\mathbf{P}^{\text{opt}}, \mathbf{F}^{\text{opt}}) = \mathbf{0}$. This concludes the proof. \square

Remark 7. The proofs show that the only assumption we need on the structure of the matrices $\mathbf{A}, \mathbf{B}, \mathbf{Q}$ and \mathbf{R} is that \mathbf{R} is block diagonal. There are no restrictions on the state, input and state cost matrices.

5 Convergence: discrete time systems

We can proceed in analogous manner as we did in the previous section. Again, we start with global convergence, where exactly the same claim holds.

Theorem 8 (Global convergence). *Let us start the split optimal policy iteration with the initial guess \mathbf{F}^0 and with updating the i^{th} subsystem first, where i is arbitrary. Then, if the pair $(\mathbf{A}^{(i)}, \mathbf{B}^{(i)})$ is controllable, then*

- (i) *all the iterates \mathbf{F}^k , $k \geq 1$, stabilize the global system;*
- (ii) *the value functions V^k corresponding to the \mathbf{F}^k satisfy $V^{k+1} \leq V^k$ pointwise; and*
- (iii) *$\mathbf{F}^k \rightarrow \mathbf{F}^{\text{opt}}$ as $k \rightarrow \infty$, where \mathbf{F}^{opt} is the optimal feedback matrix of the original problem.*

Claims (i) and (ii) of this theorem are proven the same way as for Theorem 1, the adaptation of lemmas 2–4 is done by just replacing the formulas for the continuous time system with the ones for the discrete time system. However, claim (iii) does need a little bit more work, and in order to improve the readability of the proof below, we prove the most technical part in the following lemma—which can be skipped if the reader is not interested in the details.

Lemma 9. *Let \mathbf{S} and \mathbf{R} be symmetric positive definite matrices, \mathbf{R} being also block diagonal. Set $\hat{\mathbf{I}}_i = \mathbf{I} - \Pi_i \Pi_i^T$, $\mathbf{Z} = (\mathbf{R} + \mathbf{S})^{-1}$, $\mathbf{Z}_i = \Pi_i (\Pi_i^T (\mathbf{R} + \mathbf{S}) \Pi_i)^{-1} \Pi_i^T$. Then the following equations hold:*

$$\mathbf{Z}^{-1} \mathbf{Z}_i \mathbf{Z}^{-1} - \hat{\mathbf{I}}_i \mathbf{S} \mathbf{Z}_i \mathbf{Z}^{-1} - \mathbf{Z}^{-1} \mathbf{Z}_i \hat{\mathbf{S}}_i + \hat{\mathbf{I}}_i \mathbf{S} \mathbf{Z}_i \hat{\mathbf{S}}_i = \Pi_i \Pi_i^T (\mathbf{R} + \mathbf{S}) \Pi_i \Pi_i^T \quad (10)$$

$$\mathbf{Z}^{-1} = \hat{\mathbf{I}}_i \mathbf{Z}^{-1} + \mathbf{Z}^{-1} \hat{\mathbf{I}}_i - \hat{\mathbf{I}}_i (\mathbf{R} + \mathbf{S}) \hat{\mathbf{I}}_i + \Pi_i \Pi_i^T (\mathbf{R} + \mathbf{S}) \Pi_i \Pi_i^T \quad (11)$$

Proof. To prove (10), consider the second and fourth terms on the left hand side:

$$\begin{aligned}\hat{\mathbf{I}}_i \mathbf{S} \mathbf{Z}_i \hat{\mathbf{S}}_i - \hat{\mathbf{I}}_i \mathbf{S} \mathbf{Z}_i \mathbf{Z}^{-1} &= \hat{\mathbf{I}}_i \mathbf{S} \mathbf{Z}_i (\hat{\mathbf{S}}_i - \mathbf{R} - \mathbf{S}) \\ &= -\hat{\mathbf{I}}_i \mathbf{S} \mathbf{Z}_i (\mathbf{R} + \mathbf{S}) \Pi_i \Pi_i^T\end{aligned}$$

where we used $\mathbf{Z}_i \mathbf{R} = \mathbf{Z}_i \mathbf{R} \Pi_i \Pi_i^T$, because $\mathbf{Z}_i \mathbf{R}$ has non-zero blocks only on its i^{th} block row. Analogously, the first and third terms yield:

$$\mathbf{Z}^{-1} \mathbf{Z}_i \mathbf{Z}^{-1} - \mathbf{Z}^{-1} \mathbf{Z}_i \mathbf{S} \hat{\mathbf{I}}_i = (\mathbf{R} + \mathbf{S}) \mathbf{Z}_i (\mathbf{R} + \mathbf{S}) \Pi_i \Pi_i^T.$$

Adding up these two equations gives us that the left hand side of (10) is equal to

$$(\mathbf{R} + \mathbf{S} - \hat{\mathbf{I}}_i \mathbf{S}) \mathbf{Z}_i (\mathbf{R} + \mathbf{S}) \Pi_i \Pi_i^T = \Pi_i \Pi_i^T (\mathbf{R} + \mathbf{S}) \mathbf{Z}_i (\mathbf{R} + \mathbf{S}) \Pi_i \Pi_i^T.$$

Now we note that $\Pi_i \Pi_i^T (\mathbf{R} + \mathbf{S})$ coincides with $\mathbf{R} + \mathbf{S}$ in its i^{th} block row and is zero elsewhere, and $(\mathbf{R} + \mathbf{S}) \Pi_i \Pi_i^T$ coincides with $\mathbf{R} + \mathbf{S}$ in its i^{th} block column and is zero elsewhere. The matrix \mathbf{Z}_i is the inverse of the $(i, i)^{\text{th}}$ block of $\mathbf{R} + \mathbf{S}$ in its $(i, i)^{\text{th}}$ block and is zero elsewhere. Hence the product of these matrices is exactly the right hand side of (10), which proves the equality.

To see that (11) holds, note that

- $\hat{\mathbf{I}}_i (\mathbf{R} + \mathbf{S})$ is the matrix $(\mathbf{R} + \mathbf{S})$ with its i^{th} block row set to zero,
- $(\mathbf{R} + \mathbf{S}) \hat{\mathbf{I}}_i$ is the matrix $(\mathbf{R} + \mathbf{S})$ with its i^{th} block column set to zero,
- $\hat{\mathbf{I}}_i (\mathbf{R} + \mathbf{S}) \hat{\mathbf{I}}_i$ is the matrix $(\mathbf{R} + \mathbf{S})$ with its i^{th} row and column set to zero, and
- $\Pi_i \Pi_i^T (\mathbf{R} + \mathbf{S}) \Pi_i \Pi_i^T$ is the matrix $(\mathbf{R} + \mathbf{S})$ with everything apart from its $(i, i)^{\text{th}}$ block set to zero.

Combining the terms on the right hand side as they are gives exactly the matrix $\mathbf{R} + \mathbf{S}$, i.e. \mathbf{Z}^{-1} . \square

Now everything is set up to prove Theorem 8.

Proof of Theorem 8. Statements (i) and (ii) are shown analogously as before. Statement (iii) follows from Lemma 4 (more precisely, from its discrete counterpart), if we show that any fixed point of the split optimal policy iteration (more precisely the associated matrix $\bar{\mathbf{P}}$) solves the discrete algebraic Riccati equation (5), and by uniqueness of its solutions [Son98] coincides with the optimal solution.

Since $\bar{\mathbf{P}}$ and $\bar{\mathbf{F}}$ are fixed points of the iteration, they solve (5) with the modified matrices (7) for every i . This reads as, by using the notation of the previous lemma:

$$\begin{aligned} \bar{\mathbf{P}} &= (\mathbf{A} + \mathbf{B} \hat{\mathbf{I}}_i \bar{\mathbf{F}})^T \bar{\mathbf{P}} (\mathbf{A} + \mathbf{B} \hat{\mathbf{I}}_i \bar{\mathbf{F}}) - (\mathbf{A} + \mathbf{B} \hat{\mathbf{I}}_i \bar{\mathbf{F}})^T \bar{\mathbf{P}} \mathbf{B} \mathbf{Z}_i \mathbf{B}^T \bar{\mathbf{P}} (\mathbf{A} + \mathbf{B} \hat{\mathbf{I}}_i \bar{\mathbf{F}}) + \mathbf{Q} + \bar{\mathbf{F}}^T \hat{\mathbf{I}}_i \mathbf{R} \hat{\mathbf{I}}_i \bar{\mathbf{F}} \\ &= \mathbf{A}^T \bar{\mathbf{P}} \mathbf{A} - \mathbf{A}^T \bar{\mathbf{P}} \mathbf{B} \mathbf{M} \mathbf{B}^T \bar{\mathbf{P}} \mathbf{A} + \mathbf{Q}, \end{aligned}$$

with

$$\mathbf{M} := \mathbf{Z} \left(\hat{\mathbf{I}}_i \mathbf{Z}^{-1} + \mathbf{Z}^{-1} \hat{\mathbf{I}}_i - \hat{\mathbf{I}}_i (\mathbf{R} + \mathbf{S}) \hat{\mathbf{I}}_i + \mathbf{Z}^{-1} \mathbf{Z}_i \mathbf{Z}^{-1} - \hat{\mathbf{I}}_i \mathbf{S} \mathbf{Z}_i \mathbf{Z}^{-1} - \mathbf{Z}^{-1} \mathbf{Z}_i \mathbf{S} \hat{\mathbf{I}}_i + \hat{\mathbf{I}}_i \mathbf{S} \mathbf{Z}_i \hat{\mathbf{I}}_i \right) \mathbf{Z},$$

where $\mathbf{S} = \mathbf{B}^T \bar{\mathbf{P}} \mathbf{B}$. The above equation is identical with the full discrete algebraic Riccati equation (5) if $\mathbf{M} = \mathbf{Z}$. This identity is shown in Lemma 9. Hence we have shown that the fixed point $\bar{\mathbf{P}}$ of the split optimal policy iteration also solves (5), and by uniqueness of the solutions it is identical with the solution of this full equation. This concludes the proof. \square

Having shown global convergence, we turn to the analysis of the convergence speed. Quite surprisingly, it turns out that for the discrete time case the convergence is not local quadratic in general, but is in a non-trivial relation with the coupling strength between the subsystems. The weaker the coupling between the subsystems, the faster the convergence of the split optimal policy iteration; and if no coupling is present, the convergence is locally quadratic.

Let us recall, that just as in the continuous times case, one iteration cycle can be written as $g := g^{j-1} \circ g^{j-2} \circ \dots \circ g^{j+1} \circ g^j$ (if we start the iteration at the j^{th} subsystem), where the g^i , $i = 1, \dots, n$, are the maps describing the update of the feedback matrix \mathbf{F} . Since $\mathbf{B}^{(i)}$ and $\mathbf{R}^{(i)}$ do not depend on \mathbf{F} , the map g^i can be decomposed into factors:

$$g^i : \mathbf{F} \mapsto (\mathbf{A}^{(i)}, \mathbf{Q}^{(i)}) \mapsto (\mathbf{P}^{\text{new}}, \mathbf{A}^{(i)}) \mapsto \mathbf{F}^{\text{new}},$$

where the first mapping encodes the modified system matrices due to (7), the second realizes the solution of the Riccati equation (5), and the third performs the update of the feedback matrix as described in section 3.2.

Theorem 10 (Local speed of convergence). *The split optimal policy iteration for discrete time LQR problem converges locally linearly, i.e. there is a $\varrho > 0$ such that*

$$\|\mathbf{F}^{k+n} - \mathbf{F}^{\text{opt}}\| \leq \varrho \|\mathbf{F}^k - \mathbf{F}^{\text{opt}}\|$$

The convergence rate ϱ is governed by the norm of the iteration matrix

$$Dg(\mathbf{F}^{\text{opt}}) = Dg^{j-1}(\mathbf{F}^{\text{opt}}) \cdot Dg^{j-2}(\mathbf{F}^{\text{opt}}) \dots Dg^{j+1}(\mathbf{F}^{\text{opt}}) \cdot Dg^j(\mathbf{F}^{\text{opt}}),$$

where

$$Dg^i(\mathbf{F}^{\text{opt}}) = -\Pi_i \left(\mathbf{R}^{(i)} + \mathbf{B}^{(i)\text{T}} \mathbf{P}^{\text{opt}} \mathbf{B}^{(i)} \right)^{-1} \mathbf{B}^{(i)\text{T}} \mathbf{P}^{\text{opt}} \mathbf{B} \hat{\mathbf{l}}_i. \quad (12)$$

Proof. The strategy of the proof follows that of Theorem 6, although it will turn out that $Dg^i(\mathbf{F}^{\text{opt}}) \neq \mathbf{0}$ in general. Using the above factorization of g^i we can give an expression for its derivative. We use the abbreviation $\mathbf{Z}_i = \Pi_i \left(\mathbf{R}^{(i)} + \mathbf{B}^{(i)\text{T}} \mathbf{P}^{\text{new}} \mathbf{B}^{(i)} \right)^{-1} \Pi_i^{\text{T}}$.

$$\begin{aligned} Dg^i(\mathbf{F}) &= \frac{\partial \mathbf{F}^{\text{new}}}{\partial \mathbf{F}} \\ &= \frac{\partial \mathbf{F}^{\text{new}}}{\partial \mathbf{P}^{\text{new}}} \frac{\partial \mathbf{P}^{\text{new}}}{\partial \mathbf{F}} + \frac{\partial \mathbf{F}^{\text{new}}}{\partial \mathbf{A}^{(i)}} \frac{\partial \mathbf{A}^{(i)}}{\partial \mathbf{F}} \\ &= \mathbf{Z}_i \mathbf{B}^{\text{T}} \frac{\partial \mathbf{P}^{\text{new}}}{\partial \mathbf{F}} \mathbf{B} \mathbf{Z}_i \mathbf{B}^{\text{T}} \mathbf{P}^{\text{new}} \mathbf{A}^{(i)} - \mathbf{Z}_i \mathbf{B}^{\text{T}} \frac{\partial \mathbf{P}^{\text{new}}}{\partial \mathbf{F}} \mathbf{A}^{(i)} - \mathbf{Z}_i \mathbf{B}^{\text{T}} \mathbf{P}^{\text{new}} \frac{\partial \mathbf{A}^{(i)}}{\partial \mathbf{F}} \\ &= \mathbf{Z}_i \mathbf{B}^{\text{T}} \left[\frac{\partial \mathbf{P}^{\text{new}}}{\partial \mathbf{F}} \mathbf{B} \mathbf{Z}_i \mathbf{B}^{\text{T}} \mathbf{P}^{\text{new}} \mathbf{A}^{(i)} - \frac{\partial \mathbf{P}^{\text{new}}}{\partial \mathbf{F}} \mathbf{A}^{(i)} - \mathbf{P}^{\text{new}} \frac{\partial \mathbf{A}^{(i)}}{\partial \mathbf{F}} \right] \end{aligned}$$

where the third equality follows directly from the update formula (section 3.2) and the chain rule.

Next, we are going to determine $\frac{\partial \mathbf{P}^{\text{new}}}{\partial \mathbf{F}}$ via the implicit function theorem. Note that \mathbf{P}^{new} is defined by $r^i(\mathbf{P}^{\text{new}}, \mathbf{F}) = \mathbf{0}$, where

$$r^i(\mathbf{P}, \mathbf{F}) := \mathbf{A}^{(i)\top} \mathbf{P} \mathbf{A}^{(i)} - \mathbf{A}^{(i)\top} \mathbf{P} \mathbf{B}^{(i)} \left(\mathbf{R}^{(i)} + \mathbf{B}^{(i)\top} \mathbf{P} \mathbf{B}^{(i)} \right)^{-1} \mathbf{B}^{(i)\top} \mathbf{P} \mathbf{A}^{(i)} + \mathbf{Q}^{(i)} - \mathbf{P}.$$

Let us recall that \mathbf{P}^{new} is analytic in every parameter of the Riccati equation [Del84], hence there is no obstacle in our way to use the implicit function theorem. Just as in the proof of Theorem 6 we will show that $\frac{\partial r^i}{\partial \mathbf{F}}(\mathbf{F}^{\text{opt}}) = \mathbf{0}$, implying that $\frac{\partial \mathbf{P}^{\text{new}}}{\partial \mathbf{F}}(\mathbf{F}^{\text{opt}}) = \mathbf{0}$.

To this end, we substitute in r^i the matrices from (7), differentiate it with respect to \mathbf{F} at \mathbf{F}^{opt} from (6). Then, we pull out the common factor and simplify the result by using the notation from Lemma 9 and Theorem 8. What we get is

$$\frac{\partial r^i}{\partial \mathbf{F}}(\mathbf{F}^{\text{opt}}) \cdot \Delta = \mathbf{A}^\top \mathbf{P}^{\text{opt}} \mathbf{B} \left[\hat{\mathbf{I}}_i - \mathbf{Z} \hat{\mathbf{I}}_i \mathbf{Z}^{-1} \hat{\mathbf{I}}_i - \mathbf{Z}_i \hat{\mathbf{S}}_i + \mathbf{Z} \hat{\mathbf{I}}_i \mathbf{S} \mathbf{Z}_i \hat{\mathbf{S}}_i \right] \Delta + \otimes, \quad (13)$$

where \otimes is the transposed of the first term. Consider the last two terms on the right hand side of this expression in the brackets [...]. Noting that $\mathbf{Z}^{-1} = \mathbf{R} + \mathbf{S}$, and using the shorthand \mathbf{S}_{ii} and \mathbf{R}_{ii} for the $(i, i)^{\text{th}}$ blocks of \mathbf{S} and \mathbf{R} , respectively, we obtain

$$\begin{aligned} -\mathbf{Z}_i \hat{\mathbf{S}}_i + \mathbf{Z} \hat{\mathbf{I}}_i \mathbf{S} \mathbf{Z}_i \hat{\mathbf{S}}_i &= \mathbf{Z} (\hat{\mathbf{I}}_i \mathbf{S} - \mathbf{R} - \mathbf{S}) \mathbf{Z}_i \hat{\mathbf{S}}_i \\ &= \mathbf{Z} \left(-\Pi_i \Pi_i^\top \mathbf{S} - \mathbf{R} \right) \Pi_i (\mathbf{R}_{ii} + \mathbf{S}_{ii})^{-1} \Pi_i^\top \hat{\mathbf{S}}_i \\ &= -\mathbf{Z} \Pi_i (\mathbf{S}_{ii} + \mathbf{R}_{ii}) (\mathbf{S}_{ii} + \mathbf{R}_{ii})^{-1} \Pi_i^\top \hat{\mathbf{S}}_i \\ &= -\mathbf{Z} \Pi_i \Pi_i^\top \hat{\mathbf{S}}_i, \end{aligned}$$

where we used in the third equality that $\mathbf{R} \Pi_i = \Pi_i \mathbf{R}_{ii}$ due to the fact that \mathbf{R} is block diagonal. With this, returning to the bracketed expression in (13), we have

$$\begin{aligned} [\dots] &= \left(\mathbf{I} - \mathbf{Z} \left(\hat{\mathbf{I}}_i \mathbf{Z}^{-1} + \Pi_i \Pi_i^\top \mathbf{S} \right) \right) \hat{\mathbf{I}}_i \\ &= \left(\mathbf{I} - \mathbf{Z} (\hat{\mathbf{I}}_i \mathbf{R} + \mathbf{S}) \right) \hat{\mathbf{I}}_i \\ &= \mathbf{Z} (\mathbf{R} + \mathbf{S} - \hat{\mathbf{I}}_i \mathbf{R} - \mathbf{S}) \hat{\mathbf{I}}_i \\ &= \mathbf{Z} \Pi_i \Pi_i^\top \mathbf{R} \hat{\mathbf{I}}_i \\ &= \mathbf{0}, \end{aligned}$$

since $\Pi_i \Pi_i^\top \mathbf{R} \hat{\mathbf{I}}_i = \mathbf{0}$, again, due to the block diagonal structure of \mathbf{R} .

Returning to the formula for Dg^i , only the last term in the brackets does not vanish. We know from Theorem 8 that if $\mathbf{F} = \mathbf{F}^{\text{opt}}$, then $\mathbf{P}^{\text{new}} = \mathbf{P}^{\text{opt}}$, hence we obtain

$$Dg^i(\mathbf{F}^{\text{opt}}) = -\mathbf{Z}_i \mathbf{B}^\top \mathbf{P}^{\text{opt}} \mathbf{B} \hat{\mathbf{I}}_i.$$

□

Remark 11 (Rate of convergence). *If we assume distributed actuation (i.e. that the control input of the i^{th} subsystem has only a direct influence on the i^{th} subsystem itself), the input matrix \mathbf{B} is block diagonal with blocks \mathbf{B}_{ii} . Assuming a similar block decomposition according to the subsystem structure of the other matrices involved, (12) yields that all block rows of $Dg^i(\mathbf{F}^{\text{opt}})$ are zero apart from the i^{th}*

one, and its $(i, j)^{\text{th}}$ block is

$$(\mathbf{R}_{ii} + \mathbf{B}_{ii}^T \mathbf{P}_{ii}^{\text{opt}} \mathbf{B}_{ii})^{-1} \mathbf{B}_{ii}^T \mathbf{P}_{ij}^{\text{opt}} \mathbf{B}_{jj} \quad (14)$$

for $j \neq i$, and is the zero matrix if $j = i$.

This can be interpreted in terms of coupling. If the system matrix \mathbf{A} is block diagonal as well, i.e. the whole system consists of individual non-coupled subsystems, then \mathbf{P}^{opt} is block diagonal too and the matrix (14) is zero for every j . If the system is weakly coupled, i.e. the off-diagonal blocks of \mathbf{A} are much smaller than its diagonal blocks, perturbation arguments [KPC86] show that in general this property carries over to \mathbf{P}^{opt} as well.² Thus, (14) suggests that for weakly coupled systems we should expect the split optimal policy iteration to converge rapidly because $\rho \ll 1$.

References

- [Bel57] Richard Bellman. *Dynamic Programming*. Princeton University Press, 1. edition, 1957.
- [Ber07] Dimitri P. Bertsekas. Separable dynamic programming and approximate decomposition methods. In *IEEE T. Automat. Contr.*, volume 52, pages 911–916, 2007.
- [BH02] J. Bezdek and R. Hathaway. Some notes on alternating optimization. In *Lecture Notes in Computer Science*, volume 2275, pages 288–300. 2002.
- [Bro51] G. W. Brown. Iterative solutions of games by fictitious play. In *Activity Analysis of Production and Allocation*, pages 374–376. Wiley, 1951.
- [Del84] David F. Delchamps. Analytic feedback control and the algebraic Riccati equation. *Automatic Control, IEEE Transactions on*, 29(11):1031 – 1033, nov 1984.
- [GKP01] Carlos Guestrin, Daphne Koller, and Ronald Parr. Multiagent planning with factored MDPs. In *In NIPS-14*, pages 1523–1530. The MIT Press, 2001.
- [Joh] Ramesh Johari. Lecture notes on “Dynamics and Learning in Games”. <http://www.stanford.edu/~rjohari/Teaching/LectureNotes>.
- [KJ14a] Péter Koltai and Oliver Junge. Optimal value functions for weakly coupled systems: a posteriori estimates. *ZAMM*, 94:345–355, 2014. DOI: [10.1002/zamm.201100138](https://doi.org/10.1002/zamm.201100138).
- [KJ14b] Péter Koltai and Oliver Junge. Quantized nonlinear feedback design by a split dynamic programming approach, 2014. To appear in: Proceedings of the 21st International Symposium on Mathematical Theory of Networks and Systems (MTNS 2014).
- [KPC86] M. M. Konstantinov, P. Hr. Petkov, and N. D. Christov. Perturbation analysis of the continuous and discrete matrix ricatti equations. In *Proc. 1986 ACC, Seattle, WA*, volume 1, pages 636–639, 1986.
- [Lit01] M. Littman. Value-function reinforcement learning in Markov games. *Journal of Cognitive Systems Research*, 2:55–66, 2001.
- [LR00] Martin Lauer and Martin Riedmiller. An algorithm for distributed reinforcement learning in cooperative multi-agent systems. In *In Proceedings of the Seventeenth International Conference on Machine Learning*, pages 535–542. Morgan Kaufmann, 2000.
- [MHeK⁺98] Nicolas Meuleau, Milos Hauskrecht, Kee eung Kim, Leonid Peshkin, Leslie Pack Kaelbling, Thomas Dean, and Craig Boutilier. Solving very large weakly coupled Markov decision processes. In *In Proceedings of the Fifteenth National Conference on Artificial Intelligence*, pages 165–172, 1998.
- [MS96a] D. Monderer and L. S. Shapley. Fictitious play property for games with identical interests. *Journal of Economic Theory*, 68:258–265, 1996.

²In [KJ14a] the important case is studied where weak coupling in \mathbf{A} does not carry over to weak coupling in \mathbf{P}^{opt} .

- [MS96b] D. Monderer and L. S. Shapley. Potential games. *Games and Economic Behavior*, 14:124–143, 1996.
- [OR94] Martin J. Osborne and Ariel Rubinstein. *A Course in Game Theory*. The MIT Press, 1994.
- [Rob51] J. Robinson. An iterative method of solving a game. *Annals of Mathematics*, 54:296–301, 1951.
- [Son98] Eduardo D. Sontag. *Mathematical Control Theory: Deterministic Finite Dimensional Systems*. Springer, New York, 2. edition, 1998.